Experian™
Data Quality

Data Migration Survival Guide (Part 2 of 2)

Experian Pandora

Created by Dylan Jones
Editor Data Migration Pro.com

Contents

1. Introduction

Welcome to Part 2 of the Data Migration Survival Guide. In Part 1 we focused on:

- How to spot leadership challenges (know and anticipate them all)
- There are Common Assumptions (you need to eliminate them)
- Leadership Planning starts BEFORE the migration project
- Massively de-risk your project with a PMIA

In this part of the series we look at the key stages of a data migration journey that project leaders should be requesting to increase the likelihood of success.

We go through each stage in-depth and give yout notes sections to take action and share best-practices with your team.

2. Setting up the Project Office

As a project leader, one of your first tasks is to create a firm understanding of what deliverables and project artefacts are required to deliver your migration. This may sound obvious but many project leaders rely on standard materials from their corporate project methodology of choice and unfortunately omit many resources that are distinct to migration projects.

Not only do you have to create a list of what is required but a number of templates need to be created with a corresponding workflow that outlines how each document or artefact influences the various activities and other deliverables on the project.

The benefit of this additional planning step is that everyone on the project understands:

A. What tasks they need to complete

B. What deliverables are created by each task

C. How each deliverable 'slots' into other deliverables (and tasks)

The earlier you execute this task, the smoother your project will execute.

If you are outsourcing the project entirely then it is still recommended that you work with the systems integrator to understand how they have set up their project office in terms of activities, resources and workflow. It's also useful to assess a suppliers approach during your review of the different options at the bid phase. This enabls you to establish who are the most experienced and well prepared integrators for the project.

If a supplier cannot show you examples of templates, workflow and artefacts - how can they have successfully completed a complex migration in the past?

What type of deliverables exist in the project office?

There are essentially 4 types of deliverable on a data migration project that need to be initiated in the project office.

A. Data Migration Specific

B. Corporate Specific

C. Partner Specific

D. Regulatory Specific

For example you may have corporate guidelines on security and movement of data that have to be managed. This may require audit logs to maintained and approval forms to be completed.

By far the largest category will be the Data Migration Specific materials and these may include items such as:

- Mapping Design Specification
- Testing Specification
- Data Quality Rules Process and Documentation
- System Retirement Plan
- Data Migration Project Plan
- Configuration Management Plan
- Scope and Scale Definitions
- Stakeholder Management Plan

There are more deliverables than this but it gives you some idea of what resources you will need to plan for as a project leader.

Tip: The book 'Practical Data Migration' (2nd Edition) has an excellent list of deliverables that need to be created on a data migration project.

Practical Data Migration, 2nd Edition (author John Morris)

Notes: Project Deliverables

Please use this section to outline the deliverables you need on your project

3. Landscape Analysis (LA)

Landscape analysis is the most critical activity of your entire data migration because if ignored it can spell disaster later in the project.

The goal of Landscape Analysis to:

- Fully understand your legacy data landscape
- Document how disparate records are related
- Discover where the main problem areas lie
- Assess what the likely outcome is of continuing with an assumed migration strategy

The Landscape Analysis is effectively a more detailed form of the Pre-Migration Impact Assessment (PMIA) activity that we saw in 'Data Migration Survival Guide Part 1 of 2'.

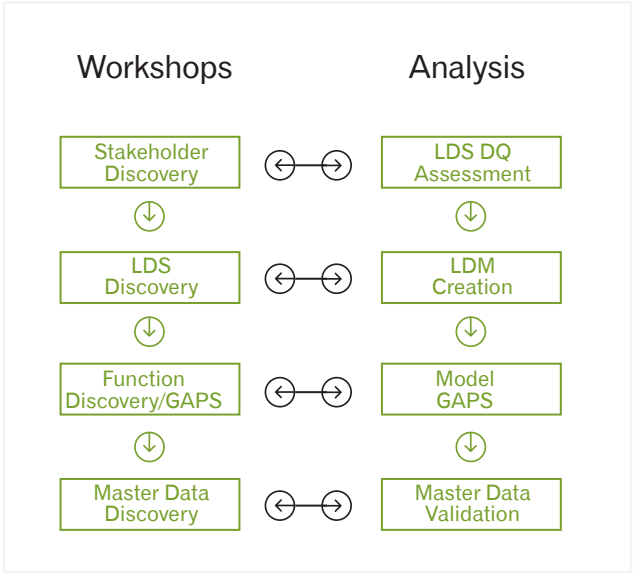
Why is Landscape Analysis so important project leaders to include?

1. Helps you dig far deeper than the earlier PMIA activity : The PMIA is typically executed over 1-4 weeks depending on the size of the project so it is limited in terms of scope. In a Landscape Analysis activity you will typically leverage a bigger team and perform much deeper data quality assessments.

2. Replaces assumptions with fact : As we saw in Part 1 of this Survival series, the biggest problem for project leaders is the sheer volume of assumptions that impact a project. Using a Landscape Analysis process can eliminate all of these assumptions and replace them with managed activities and facts.

3. Gets your team productive from day 1 : There is often a 'down tools' attitude on data migration projects before the target system is ready. A Landscape Analysis activity gets the whole team actively engaged in solving problems and creating vital project deliverables much quicker.

4. Fills in the knowledge gaps from years of misuse and poor documentation : Your legacy systems will have a great deal of missing information relating to functional and data design changes. Also, the data quality levels will probably be unknown so a Landscape Analysis is vital for filling in all the missing documentation by using data discovery, profiling and various workshops.



Landscape Analysis integrates tool based analysis with workshops involving various stakeholders and user groups.

On the tool side, you will need to use advanced Data Discovery, Data Profiling and Data Quality Management tools to perform Landscape Analysis fully.

Did you know? Experian Pandora Software provides a complete solution for supporting the Landscape Analysis activity and the full data migration.

To arrange a demonstration please contact us or visit: http://www.edq.com/uk

A key technical requirement is the relationship discovery function. Within your legacy landscape you will often find a number of disparate systems, spreadsheets, local databases and a myriad of other data sources. Having a relationship discovery function allows you to quickly map all of these sources together, irrespective of data formats.

What other tools are useful during Landscape Analysis?

Logical and Physical Modelling Case Tools: Where there are no historical models of the legacy environment these tools can be particularly useful. They are also very useful for identifying obvious gaps between legacy models and the target model.

Business Glossary and Data Dictionary : One of the key challenges of a migration is ensuring that all of your legacy and target data meanings are fully understood, shared and governed. By adding a Business Glossary or Data Dictionary to your project you can create one source of knowledge for the entire project. Note that some tools combine a Business Glossary with the data quality product itself.

Notes: Landscape Analysis

Please use this section to outline your approach to Landscape Analysis and where any changes need to be made

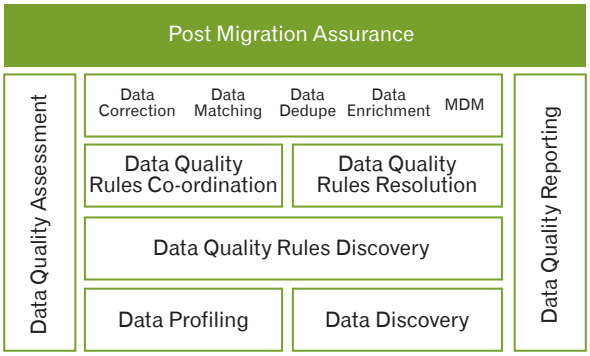
4. Data Quality Management (DQM)

The DQM capability of your project is incredibly important and often entirely overlooked by project leaders in a bid to cut costs and perceived effort.

Your goal as a project leader is to deliver a zero defect data migration. This is the central mission of the entire project and defines your approach to all tasks. Without Data Quality Management throughout the entire project you simply cannot achieve a zero defect migration.

Your task as project leader is to ensure that your project has the right data quality skills, technology and methodology. Without these, your chances of success are severely limited.

If you are outsourcing project, either in part or completely, then you still need to ensure that the right data quality management approach is being adopted and this should form part of your benchmark review of suppliers at the start of the project.



You need to assess your current capabilities agains the functions listed in the above diagram.

Note that several of these functions may not be required depending on the type of data and nature of the project. For example, you may not need data deduplication functions if no duplicates exist in your data but this is rare, there is generally a large amount of duplication in legacy data sets.

Data Quality - Leadership Duties

Project Sponsor :

- Ensures resources are made available
- Reviews regular updates

Project Leader :

- Implements processes and procedures
- Ensures team are fully trained and tooled up
- Reviews progress, manage politics

Team Leader :

- Ensures appropriate attendance and training
- Commits resources to data quality functions

Data Quality Rules

Managing data quality rules is the cornerstone of your data migration project. Without this activity you will have little chance of delivering an on-time, on-budget migration.

The purpose of creating data quality rules is to ensure that your legacy data is 'fit for purpose' on its journey from the source systems, through the migration architecture, into the target system and most importantly, into the functional uses required of the business users.

The Landscape Analysis activity is one of the main sources of intelligence gathering for the data quality rules process. All of the data profiling and data discovery activities will create a large volume of data quality rules that need to be managed.

Many people assume that you only create data quality rules for defective data, this isn't the case. It is advisable to create data quality rule specifications for all your data so that these rules can be managed by specialist data quality software that monitors and reports on the quality of the data leading up to the migration. Using this approach ensures that you have zero defects at run-time migration, which is your primary goal.

The 'Perfect Data' Illusion

Some people will comment that the legacy data is faultless, there are no reported errors. This may in fact be true but can be a red herring as your legacy data has to embark on a journey to a new environment that will place greater pressures on your source data, pressures that in many cases it was never designed for.

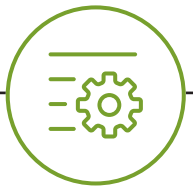
Never assume that just because no-one has reported source data defects in the past it will still be fine to migrate the data 'as-is'. This is why data quality rules are so important, they replace assumptions with cold, hard facts.

For a more detailed account of Data Quality Rules, you may find these resources useful:

- How to Create a Data Quality Rules Process for Data Migration, with John Morris
- Data Quality Rules: The Definitive Guide to Getting Started

Notes: Data Quality Management

Please use this section to outline your approach to Data Quality Management and where any changes need to be made



5. Agile Design & Build

Now we move on to the physical design and construction of your architecture that will successfully migrate the data.

As a project leader you may think this is a largely technical endeavour but you do have a role to play as you will help determine how the project is structured and delivered.

The classic data migration design and build strategy is a waterfall approach, typically along the lines of:

1. Analyse
2. Design
3. Build
4. Test
5. Execute

The final go-live execution phase typically requires a 'Big Bang' style migration where all the data is shunted across in one movement, typically over a long-weekend or other period of agreed downtime within the source system.

There are some problems with this approach however:

- Long projects mean requirements can change
- Business waits a long time for any sign of progress
- Risky to deliver in a tight timeframe
- Legacy data and systems are constantly changing so bad data can creep in

One emerging trend in data migration projects is to take a more agile approach to project management. Instead of trying to tackle the entire project one phase at a time you create cycles of delivery that give the business something tangible, typically in a much shorter timeframe.

Here are some of the benefits of Agile Data Migration:

- Focus is on shorter delivery cycles, faster released value
- Risks are reduced because migration is broken into smaller chunks
- Business is not overwhelmed with a huge change program
- Not reliant on a long window of opportunity for migration
- Big Bang can be the end goal but also supports phased implementation

There are challenges however. Firstly, you need to have a team who understands the Agile mindset and as a project leader you need to have the confidence to manage a project within the Agile discipline.

Secondly, you have to understand that if you are delivering smaller 'chunks' of data to the target environment perhaps over several months, instead of just one big bang, then you need to also manage the coordination of business user migration.
For example, using the Agile approach, some migrations are executed by specific subsets of data e.g. geography (country/region/continent), product type, customer type etc.

For more thoughts on Agile Data Migration, also read:

Data migration - the agile experience
Working with Suppliers - Agile considerations

One of the big obstacles to overcome for the project leader interested in following an Agile approach is integrating suppliers into your project who may not be used to the Agile way of working.

Here are some questions to consider:

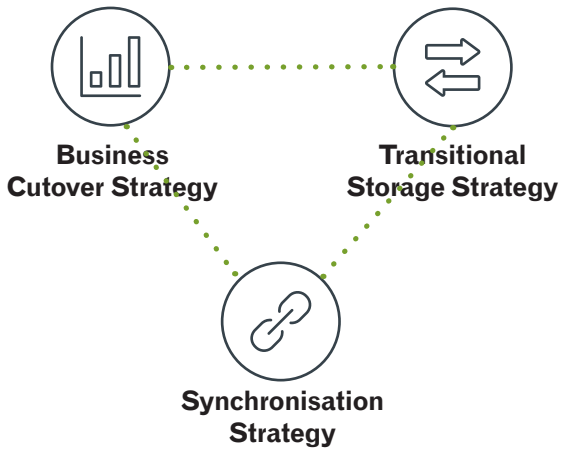
- What is the 'Unit of Migration' and how it will impact the target system?
- How will we manage data quality issues in a more agile way?
- What sprint durations will we set and what is the reasoning for setting them at this level?
- Are business and target system owners fully aware of their obligations if we go down an Agile and Phased approach?

Notes: Agile Design and Build

Please use this section to outline any actions you need to take following the Agile Design and Build section.

6. Go-Live Migration Strategy

The Go-Live strategy is essentially 3 strategies rolled into 1 and this obviously needs a great deal of direction from the project leader to manage correctly.



Business Cutover Strategy

This strategy outlines how the business will be transitioned from the legacy (source) environment to the new target environment.

Big Bang Cutover : As we discussed earlier, the most common approach is 'Big Bang'. Business usage on the source systems ceases and all of the services switch to the target system following the successful migration of data.

Incremental Cutover : Using this approach, the business is moved across incrementally, in a phased transition. For example, a bank may move all savings accounts first, followed by all current accounts in a second increment. The business services associated with each dataset would move at the same time so that savings functionality is supported on the target system until the savings accounts are transported at a later date.

Parallel Cutover : In this approach, the business maintains two environments both source and target, simultaneously. All of the data has been migrated but the business may decide to keep both environments active so that they can revert back to the source environment where required.

Transitional Storage Strategy

This is more of an architectural design issue but still something the project leader should be interested in as it may impact budget.

When we migrate data, the information has to be stored somewhere on its path from source to target. Some migration tools (such as Experian Pandora) can hold all of the data in its own repository so there is no need for additional storage requirements.

Other solutions require the use of what is referred to as a 'Staging Area'. This is kind of a half-way house architecture that more closely resembles the target system. Data is extracted from the source environment, transformed into the target specifcation and then stored in the staging area. It may require further processing in the staging area (e.g. to cleanse or correct known problems. Eventually the data is loaded into the target environment either from the staging area, or as in the case of Experian Pandora, straight from its own repository instead.

Another option is where the data is migrated in one active operation, straight from the source environment, through the migration controller and finally into the target system. The data is stored in-memory during the migration.

So, 3 different options for the project leader to consider and discuss with their technical team.

Synchronisation Strategy

Depending on the type of migration architecture and Business Cutover strategy you adopt there may be a need to keep the data between the source and target environments in sync.

For example, a parallel approach may require customer bank accounts to be maintained in both source and target systems following the migration. This is carried out to ensure that the business is happy with the target system functionality. In this scenario you will need a mechanism to synchronise changes in each system so that both systems always have accurate information.

Another strategy may be to migrate all of the data to a target system but maintain service only from the source system. The target system can be synchronised with any changes in the legacy world until such time that the business is happy for the target system to go fully operational.

There are essentially two synchronisation strategies that can be adopted:

- Option 1: Mono-Directional Sync
- Option 2: Bi-Directional Sync

Mono-Directional is where changes in one system are only passed to the other system and not vice-versa.

Bi-Directional is where both systems can pass or receive changes.

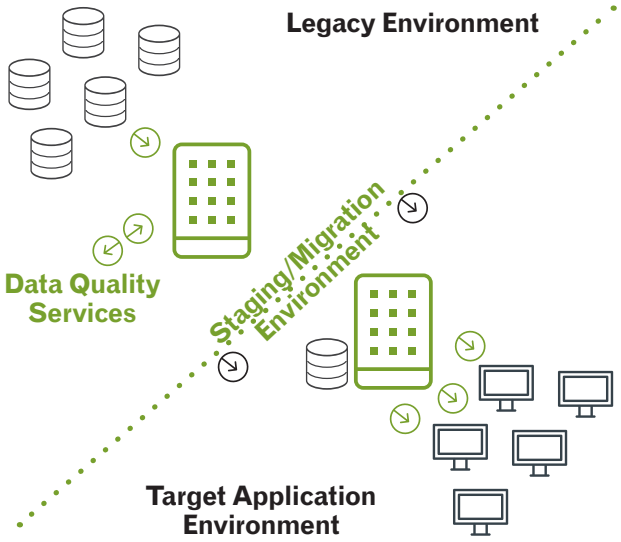
- Go-Live Migration : Important Leadership Considerations
- What are the relative costs vs risks vs benefits of each tactic?
- Are we skilled enough to pull off anything complex?
- Can our technology cope with complicated sync options?
- Will the target interface perfomance cancel the benefit of high-performing ETL tools?
- Is our fallback strategy independently validated and tested?
- How will the business support a parallel run?
- Can we afford to wait for a one-off, Big Bang implementation?
- Have we got expert advice available if required?

Notes: Go-Live Migration Strategy

Please outline any actions you need to take following the Go-Live Strategy section.

7. Testing and Assurance

Testing is obviously a critical activity for the project leader to schedule and organise. Many projects fail through lack of effective testing strategy so this is certainly something you want to plan well in advance.



The above diagram outlines the typical scope of a data migration and where testing needs to be applied. As a project leader you have to determine the scope of your teams influence.

Data Quality Management vs Data Migration Testing

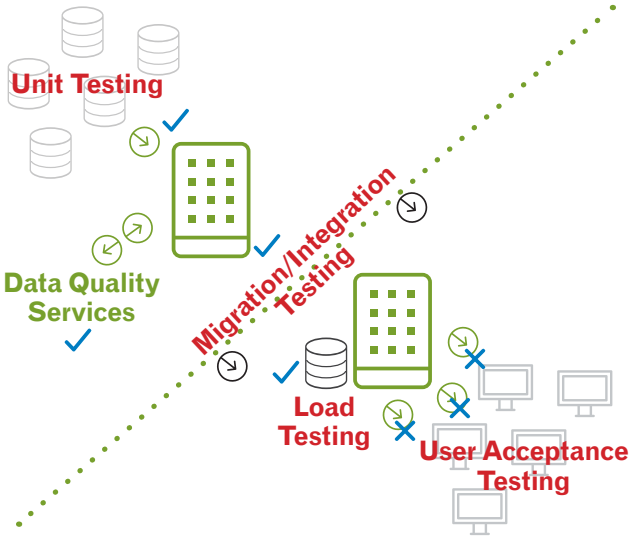
There is often some confusion around the difference between data quality management and testing a data migration.

Data Quality Management:

- Validates if data is fit for different stages of the migration
- Data Quality Rules can be used in Testing scripts
- Defects are typically fixed before the migration, often by the business

Data Migration Testing:

- Hierarchy of tests to ensure the entire data migration will succeed
- Starts with small unit tests, ends with full dry-run testing
- Defects are typically fixed before the migration, often by technicians



The above image outlines where the data migration testing activities will typically take place.

8. Archival and Decommissioning

If you're not decommissioning your legacy system following the migration then technically, you're not performing a data migration.

The aim of every migration is to ensure the seamless transition of data from source to target systems but also the orchestrated shut down of the legacy systems that no longer have a function.

Many business leaders struggle with this decision but it's often because they were not bought into the process right from the outset. They get scared of a future that doesn't include the legacy systems they know and trust.

As a data migration project leader it is your job to allay those fears and have an active plan in place to decommission the legacy systems.

Some of the activities you will have to be mindful of include:

- The Decommission Plan must include a fallback plan in event of target migration failure
- If outsourcing the migration, your role is to check the viability of fallback plan and ensure
- Client-side leadership role is to ensure stakeholders are committed to decommission schedule
- A 'System Retirement Plan' is a sponsors binding commitment to decommissioning

Here are some further resources to help with this topic:

Does your data migration strategy have an effective plan for systems retirement?

What should you include in a System Retirement Plan?

- Step-by-step instructions for system shutdown
- Action plans of any transitional processes that the business users must enact e.g. dual data entry
- Specification of precise testing regime that will satisfy all data migration stakeholders

For some industries there are data protection implications for archiving data so do check the applicable policies for your region. For example, there may be data retention laws surrounding customer data (e.g. Data Protection Act).

In some industries there needs to be a full audit trail of the entire migration (e.g. Pharma regulations in the USA) that may mandate some form of legacy data retention.

Also, be wary of security policies (e.g. encryption of medical records, passwords, credit card details, personnel records) when archiving your data.

Some products, such as Experian Pandora, enable the archival of data directly from the migration platform itself for storage in an active repository where historical data can be accessed.

9. Post-Migration Data Quality Assurance

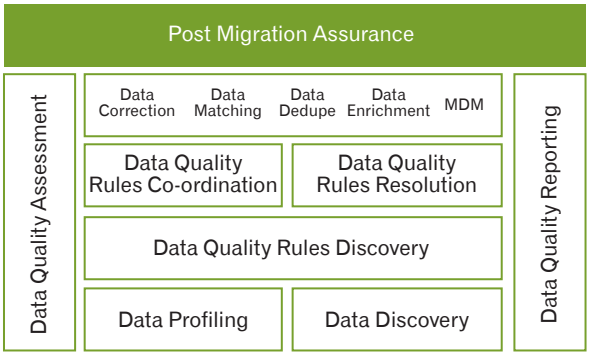
One of the biggest mistakes project leaders make is to waste the valuable resources that have been created on their migration.

Consider some of the data quality resources that are created on a migration:

- Data Quality Rules Process
- Technology, training and governance
- Data Quality uplift
- Zero defect target environment

The question that must be raised is why invest in protecting the data assets of the target environment only to see it be lost in the ensuing weeks and months after the migration has completed.

If you consider the skills, deliverables and resources that you have created earlier in the project then it makes a great deal of sense to create an ongoing data quality assurance capability to protect the quality of data moving forward.



As the diagram above illustrates, through the data migration process you actually create a fully operational data quality capability. With additional funding this capability can live on after the migration and even begin to extend its reach, supporting interconnecting systems and business functions across the organisation.

Notes: Testing, Archival andOngoing Assurance

Please outline any actions you need to take from the last 3 sections of this guide.

10. Do You Need Help?

If you have any questions about the topics raised or how to apply next generation data quality and data migration solutions on your project then please contact us:

E: **dataquality@experian.com**
T: **0800 197 7920**

Dylan Jones
Editor, Data Quality Pro

Dylan Jones is the Editor and Founder of Data Quality Pro and Data Migration Pro. He has over 20 years' experience of delivering complex data quality and data-driven initiatives. Dylan is a keynote speaker, author and regular publisher of expert insights on data quality related topics.

About Experian Data Quality

Experian Data Quality has built up exceptional market coverage assisting customers with their unique data quality challenges. We provide a comprehensive toolkit for data quality projects combining our market leading software with a vast scope of reference data assets and services. Our mission is to put our customers in a position to make the right decisions from accurate and reliable data. The size and scope of data management projects varies considerably but the common factor in all ventures is unlocking operational efficiency and improving customer engagement. We see the potential of data. Whether it's in enabling ambulances to be sent to the exact location of an emergency or attributing charitable donations to the people who need it the most - data accuracy makes all the difference to service provision.

Experian Australia Pty Ltd
Level 6, 549 St Kilda Road
Melbourne, VIC 3004, Australia

T (61) 3 8699 0100 | F (61) 3 9923 6280
E info@au.experian.com | W edq.com/au

Experian™
Data Quality

Intelligent interactions.
Every time.